DOCUMENT RESUME

ED 354 369 CE 063 094

AUTHOR Holburn, P. T.
TITLE Recommendations for Fair and Unbiased Apprentice Selection. An NTB/HRSC Report. Contract Report.
INSTITUTION Human Sciences Research Council, Pretoria (South Africa).
REPORT NO C/PERS-CR-454; ISBN-0-7969-1286-6
PUB DATE 92
NOTE 54p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adult Education; *Apprenticeships; Culture Fair Tests; Educational Research; Employer Attitudes; *Evaluation Methods; Foreign Countries; National Surveys; *Personnel Selection; Postsecondary Education; *Social Bias; *Test Bias; Testing Problems; *Test Selection; Vocational Education
IDENTIFIERS *South Africa

ABSTRACT

This document presents results of research to investigate bias and fairness in apprentice selection procedures used in South Africa. Chapter 1 discusses the aim of the project and summarizes the report's contents. Chapter 2 presents a brief overview of the contents of five interim reports. "Investigation of Apprentice Evaluation Procedures with Special Reference to Test Bias and Fairness" (J. Watt) documents findings from interviews with employers and trainers of apprentices. "Test Bias" (T. Tyler) discusses various types of test bias, evaluates techniques available to detect the kinds of bias, and proposes a procedure to detect biased test items. "Apprentice Selection" (P. Holburn) reports findings of a nationwide survey of apprentice selection procedures and policies. "Selection Decisions" (P. Holburn), a review of the literature on fairness in selection, discusses and evaluates models of fair selection and makes recommendations for fair selection. "Test Bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox, and High Level Figure Classification Tests" (P. Holburn) describes results of the test bias analysis of the four tests most commonly used for apprentice selection in South Africa. Chapter 3 makes recommendations for fair and unbiased apprentice selection. These different selection methods are considered first: recruitment, application forms, interviews, psychometric testing, psychomotor testing, and trainability testing. Guidelines for the selection procedure are then discussed. (Contains 17 references.) (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED354369

Contract Report C/PERS 454

RECOMMENDATIONS FOR FAIR AND UNBIASED APPRENTICE SELECTION

an NTB/HSRC report

P T Holburn

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CE063094

BEST COPY AVAILABLE

Contract Report C/PERS 454

RECOMMENDATIONS FOR FAIR AND UNBIASED APPRENTICE SELECTION

An NTB/HSRC report

P T HOLBURN

Pretoria
Human Sciences Research Council
1992

P T Holburn MSc. Senior Researcher

Group: Human Resources
General Manager: Dr K F Mauer

ISBN 0 7969 1286 6

© Human Sciences Research Council, 1992

Printed and published by the HSRC
134 Pretorius Street, Pretoria.

Acknowledgements

Dr K F Mauer: General Manager, Human Resources

Mr R F Skawran: Manager, Human Development

Dr T R Taylor: Head, Cognitive and Personality Research Programme

National Training Board

KnowledgeTec, Johannesburg

Jane England

Ricka Richter

Special thanks are due to Dr T R Taylor for the suggestions and comments he made when reviewing earlier drafts of this report.

ABSTRACT

In 1986 the National Training Board, in response to the increasing numbers of apprentices from all races, commissioned the NIPR to investigate bias and fairness in apprentice selection procedures.

During the course of this research five interim reports were published as various stages of this project were completed. In this document, the final report, the contents of the previous five interim reports are summarised, and findings pertaining to, and recommendations for, fair and unbiased apprentice selection are discussed.

EKSERP

Die Nasionale Opleidingsraad het in 1986, in reaksie op die toenemende getalle vakleerlinge afkomstig uit al die rasse, die NIPN opdrag gegee om sydigheid en regverdigheid in die keuringsprosedures vir vakleerlinge te ondersoek.

Oor die tydperk wat die navorsing geduur het is vyf interim verslae gepubliseer soos die verskillende stadia van die projek voltooi is. In hierdie dokument, wat die finale verslag is, word die inhoud van die vorige vyf verslae opgesom, en bevindinge met betrekking tot en aanbevelings vir regverdige en onsydige keuring van vakleerlinge word bespreek.

CONTENTS

1. **AIM OF THE PROJECT** 1

2. **SUMMARY OF PAST REPORTS** 5

2.1 **Investigation of apprentice evaluation procedures with special reference to test bias and fairness: Pilot study - progress report - J Watt** 5

2.1.1 *Introduction* 5

2.1.2 *Objective of the study* 5

2.1.3 *Methodology* 5

2.1.4 *Conclusions* 6

2.2 **Test bias: The roles and responsibilities of test user and test publisher - T Taylor** 8

2.2.1 *Introduction* 8

2.2.2 *Objective of the study* 8

2.2.3 *Methodology* 9

2.2.4 *Conclusions* 9

2.3 **Apprentice selection: An HSRC/NTB survey of policies and methods used in the RSA with an emphasis on psychometric testing - P Holburn** 13

2.3.1 *Introduction* 13

2.3.2 *Objective of the study* 13

2.3.3 *Methodology* 14

2.3.4 *Conclusions* 14

2.4 **Selection decisions: The quest for fairness - P Holburn** 17

2.4.1 *Introduction* 17

2.4.2 *Objective of the study* 18

2.4.3 *Methodology* 18

2.4.4 *Conclusions* 18

2.5 **Test bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox and High Level Figure Classification tests - P Holburn** 23
2.5.1 *Introduction* 23
2.5.2 *Objective of the study* 23
2.5.3 *Methodology* 23
2.5.4 *Conclusions* 24
3. **CONCLUSIONS AND RECOMMENDATIONS** 27
3.1 **Factors to be taken into account when making selection decisions** 27
3.2 **Guidelines for implementing fair selection** 29
3.2.1 *Selection methods* 29
3.2.1.1 Recruitment 30
3.2.1.2 Application forms 30
3.2.1.3 Interviews 31
3.2.1.4 Psychometric testing 32
3.2.1.5 Psychomotor testing 35
3.2.1.6 Trainability testing 36
3.2.2 *The selection procedure* 37
REFERENCES 43

1. AIM OF THE PROJECT

In this country at this time, two issues relating to manpower are particularly important. The first concerns the availability and utilisation of skilled technical people and the second, the fairness of employment practices.

In an economy that is functioning productively jobs will be performed by effective, well-trained individuals. Although it is difficult and probably erroneous to single out any particular occupational group as more important than another, it is generally recognised that in a developing country such as South Africa the role of skilled technical manpower is critical. Engineers, technicians and artisans are personnel whose contribution to a successful economy is seen as vital[1]. In recognition of the importance of skilled manpower many companies sponsor the training of applicants for technical jobs. Bursaries are often available for suitable candidates wishing to study engineering, and job applicants who want to become artisans may be eligible for apprenticeships.

For all jobs selection is important. It is simply not feasible to place any job candidate in any position. For a job to be performed effectively the job incumbent must have the necessary experience and ability. When a company intends spending large sums of money on the training of an employee, it becomes extremely important that the applicants selected be successful. Candidates who are unlikely to fail or drop out of their training course must be selected. Many organisations indenture a prospective artisan as an apprentice for a number of years and support the training of the apprentice until he or she qualifies as an artisan. Clearly the expense involved in such an undertaking can be considerable; hence, the company needs to be fairly sure that only apprentices who are likely to succeed are selected. Thus effective apprentice selection is a very real concern for many organisations.

Alongside the concern with the economy and skilled technical manpower, the fairness

---

[1] Although we are at present in a recession and the demand for engineers and artisans has been drastically reduced, in times of economic growth, and *for* economic growth, technical manpower is necessary.

of traditional apprentice selection methods for individuals from all cultural groups is currently receiving a great deal of attention.

Until fairly recently past legislation in the employment and educational spheres in South Africa had the effect of limiting the candidates for most types of jobs to a single population group. This situation is changing rapidly. Many types of skilled work, for example artisan work, are now done by members of all population groups. These changes in employment patterns have important ramifications for selection.

Psychometric testing is often performed as part of apprentice selection and it sometimes happens that an applicant is rejected or accepted on the basis of a test score. In circumstances such as these when a test score can have such a profound influence on the job applicant's prospects, it is imperative that these instruments be valid for all applicants regardless of their cultural background. Because apprenticeships were previously open to white applicants only, many of the tests used for apprentice selection were developed for and standardised on one race group. Consequently, as an examination of many test manuals will reveal, information concerning test scores is available for white applicants only. Any comparison of test results between members from different race groups becomes problematic as it cannot be assumed that scores have the same meaning for different groups. Although it is possible that traditional apprentice selection methods may still be appropriate, the usefulness and validity of these methods has to be investigated for all groups to enable employers to make correct and fair decisions about all applicants.

An examination of overseas literature reveals that of all the selection methods, the use of psychometric tests is the most contentious method when multicultural selection is conducted. Numerous complaints can be heard from many quarters that tests are culturally biased. However, despite the complaints registered against tests and testing, research has shown that tests have higher validity than most other selection instruments, for example interviews (Hunter & Hunter, 1984; Schmidt & Hunter, 1986) and should therefore not be easily discarded. In fact the only selection method that sometimes has slightly higher validity than psychometric ability tests is work sample

tests. Apart from validity issues, the alternatives to tests are not necessarily less biased and it should be kept in mind that when bias is present in selection instruments it is easier to identify and eradicate from tests.

Selection, as with many other employment practices, is influenced by the social, political and economic environment in which the company operates and these factors have to be incorporated in selection decision making. In the past, legislation was enacted which prevented certain categories of people from entering certain occupations. It is not impossible that in the foreseeable future, legislation may require employers to adhere to selection practices that do not discriminate unfairly against different groups. The fairness of apprentice selection procedures is therefore a matter of genuine concern at the moment.

In a joint Human Sciences Research Council and National Training Board document (HSRC, 1985) examining artisan training in South Africa, the possibility of cultural bias in apprentice selection was noted and it was proposed that selection instruments should be examined for bias. Consequently, five years ago the National Training Board contracted the NIPR to conduct a five year project investigating apprentice selection procedures. The title of the project was "Investigation of apprentice selection procedures with special reference to test bias and fairness". The aim of this research was to investigate apprentice selection methods currently used in South Africa, to examine cultural bias in the psychometric tests used for apprentice selection and to make recommendations for fair and unbiased apprentice selection.

During the course of this project five interim reports were submitted as various stages of the project were completed. The five reports dealt with the following issues:

1. A pilot study was conducted and the findings reported in the publication "*Investigation of apprentice evaluation procedures with special reference to test bias and fairness: Pilot study - progress report*", by J Watt (1987). In this document the findings from interviews with employers and trainers of apprentices were documented. This study enabled the researchers involved with the project to

obtain first hand knowledge of apprentice selection and training and the associated problems experienced in industry.

2. A second interim report by T Taylor (1987) contained the results of an overview of the literature on test bias and was published as "*Test bias: The roles and responsibilities of test user and test publisher*". In this publication the various types of test bias were discussed and the techniques available to detect the kinds of bias evaluated. A procedure to detect biased test items was proposed.

3. A nation-wide survey of apprentice selection procedures and policies relating to apprentice selection in South Africa was undertaken. The findings from this survey are contained in a report written by P Holburn (1989) entitled "*Apprentice selection: An HSRC/NTB survey of policies and methods used in the RSA with an emphasis on psychometric testing*".

4. A review of the literature on fairness in selection was carried out by P Holburn (1991) and summarised in a publication "*Selection decisions: The quest for fairness. An HSRC/NTB report*". In this report a number of models of fair selection were discussed and evaluated. Recommendations for fair selection were made.

5. The test bias techniques discussed in the second report were applied to the four tests most commonly used for apprentice selection in South Africa. A report describing the results of the test bias analyses entitled "*Test bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox and High Level Figure Classification tests*" was written by P Holburn (in press).

This document is the final report of this project and has as its aim the presentation of guidelines for fair apprentice selection. In this document the findings of the previous interim reports are summarised and recommendations for fair apprentice selection made.

2. SUMMARY OF PAST REPORTS

In this chapter a brief overview of the contents of each of the interim reports is presented and some discussion entered into.

2.1 Investigation of apprentice evaluation procedures with special reference to test bias and fairness: Pilot study - Progress report J Watt

2.1.1 Introduction

The primary reason for conducting research into selection practices is the importance of effective selection to the organisation. Effective selection procedures help to ensure the economic viability of the organisation as they reduce training time, increase productivity and may reduce turnover.

2.1.2 Objective of the study

A review of the literature on apprentice selection can only provide a partial picture of the topics requiring attention and will not provide sufficient information on the unique aspects of the South African situation. A pilot study was thus undertaken to identify the specific problems experienced by organisations in South Africa. The aim of the pilot study was to enable the researchers to familiarise themselves with apprentice selection from the perspective of industry.

2.1.3 Methodology

The researcher (J Watt) personally conducted unstructured informal interviews with many different people involved in apprentice selection. Included in the interviews were apprentice employers, trade union representatives, technical college staff, Department of Manpower and National Training Board members and personnel from apprentice training centres. Open ended questions were used to elicit information in the following subject areas:

- Recruitment of apprentices
- Selection procedures:
    - Application forms
    - Psychometric tests
    - Trainability tests
    - Psychomotor tests
    - Interviewing
- The impact of technology and the recession on the selection and training of apprentices.
- Selection and modular training
- Industry-specific issues with respect to selection

2.1.4 Conclusions

Many of the findings from the interviews were similar to those obtained from the nation-wide survey and are therefore discussed in section 2.3. However, several findings and comments from this pilot study are particularly noteworthy and are highlighted below.

Watt noted that application forms tend to be used as a preselection instrument, particularly in a recession when many applicants are available. The biographical information supplied by a candidate can be a major determinant of whether or not he or she will be appointed. An examination of the validity of the information requested in application forms therefore seems necessary as it is clearly not fair to reject an applicant on the basis of biographical details if these have no relation to apprentice performance.

One piece of information which employers felt was particularly important is the highest school standard passed. A combination of many available applicants and technological advancement was proposed as the reason for many employers insisting that prospective apprentices have Std 10 with Maths and Science. It was noted that for the electrotechnical trades in particular, which had experienced rapid recent technological

development, many employers stated that a candidate should have at least a Std 10 certificate. In her report Watt pointed out that the insistence on higher schooling entrance requirements could result in fewer potential black apprentices as proportionally fewer blacks than whites have a Std 10.

At the time the interviews were conducted South Africa was experiencing a recession and the recruitment of apprentices had consequently decreased. In such circumstances, where little active recruitment of apprentice applicants is undertaken, the chances of increased black representation in the company are diminished. Although many organisations do not recruit new personnel during a recession, they sometimes upgrade or redeploy present staff. It is feasible that in this way more semi-skilled black workers could receive the opportunity to be trained as apprentices.

Virtually no selection is undertaken without an interview. Most employers will not accept an apprentice without having interviewed the applicant. Because the interview is likely to remain as a method of selection, Watt recommended that consideration be given to developing a standard interview form and to validating interviewing techniques.

Various forms of testing are often conducted during selection and include psychometric, trainability, work sample and psychomotor testing. Of all the forms of testing for apprentice selection, psychometric testing is the most widely conducted, with a few employers utilising trainability or job sample tests. Trainability testing has been found to be equally as valid as psychometric testing and sometimes the correlations between test and criterion scores are higher than those with conventional psychometric tests. In addition to having better validity, trainability testing is usually perceived as being fairer than other forms of testing because of the obvious relationship between the test and the job. Watt noted that several apprentice employers stated that they would like selection instruments to measure "hands-on" skills and abilities rather than purely cognitive functioning. Because trainability testing is possibly more valid than conventional testing and is perceived as being fairer and more relevant than psychometric testing, Watt recommended that for apprentice

selection trainability testing be combined with conventional psychometric testing. She noted that modular training could be used to facilitate trainability testing as the content of training modules could provide material for the design of trainability tests.

From the interviews conducted it was clear that technological development had influenced apprentice selection. The cognitive demands on apprentices in trades affected by technological advancement were much higher than they had been in the past. Higher test scores, more years of formal schooling and continued apprentice training were considered necessary to cope with technological development.

Watt reported that in addition to technological sophistication, the size of the organisation was the other factor that appeared to affect apprentice selection substantially. In smaller companies selection tended to be very informally conducted, whereas the approach in larger organisations was more thorough and professional. Several larger organisations had also considered developing and implementing specific fair selection policies such as using broadly-defined quotas.

2.2 **Test bias: The roles and responsibilities of test user and test publisher**

T Taylor

2.2.1 Introduction

Many South African psychological tests used for selection were developed for and standardised on a single racial group, for example whites. Because the job applicants for positions in the middle and lower-middle job levels are increasingly drawn from all racial groups, the suitability of tests for all examinees is currently being questioned. If tests are to be used fairly it is necessary to investigate the comparability of test scores for people from different backgrounds.

2.2.2 Objective of the study

The main purpose of this report was to evaluate item bias detection techniques and

to suggest a strategy to minimise the number of items falsely identified as biased and falsely identified as unbiased. The roles and responsibilities of test constructors and test users were examined and guidelines presented for the development of procedures which the test constructor should apply in order to identify item bias in tests.

2.2.3 Methodology

This report was in the form of a literature overview. Articles on test bias, mostly emanating from the USA and Europe, were summarised and critically evaluated. In particular item and predictive bias were discussed and strategies for their detection proposed.

2.2.4 Conclusions

As suggested by the title, in this report Taylor discussed the roles and responsibilities of the test constructor and test user. The main function of the test constructor was noted to be the examination of tests for item bias. Item bias detection procedures can easily be applied during the test development process and any items that emerge as biased replaced with unbiased ones. However, the examination of predictive bias is, according to Taylor, largely the responsibility of the test user. Because a test can be used for many different purposes it is only possible for the test constructor to examine predictive bias in a few of the more common situations.

Test bias is a controversial topic and many individuals may claim, for a variety of reasons, that a particular test (or test item) is biased. In this report Taylor noted that it has been suggested that it may not be concluded that a test is biased simply because different groups have different average scores, because so-called experts have decided on the basis of a subjective examination of the test items that the test is biased or that the test was developed for one group only. Although test items should not be stereotypical or offensive to any group, an investigation of test bias usually means that statistical analyses will be necessary.

Because the two terms "bias" and "fairness" are often confused, Taylor pointed out that bias and fairness are not the same concepts. Fairness is related to the use of tests and it is only when decisions are made on the basis of test scores that one can speak of the fairness of these decisions. Fairness models are based on values. On the other hand, bias in a test or test item is determined by applying statistical formulas. For this reason, although there is likely to be agreement on whether bias is present or absent in a test, there is no single fair way to use test scores.

In this report predictive bias and item bias were discussed. These two concepts were defined and the techniques used to detect bias presented. In this research project item bias was accorded greater prominence.

There are many different ways to detect item bias. The various methods of detecting item bias were presented, research concerning the effectiveness of the different methods was evaluated and a strategy to detect item bias was proposed.

Taylor noted that two types of item bias have been defined - unconditional and conditional. The unconditional definition is as follows:

> an item of a test is biased against (or for) members of "minority" Group B if, on that item, the members of this group obtain an average score which differs from the average score of "majority" Group A by more (or less) than expected from Group B's performance on other items of the test. (p. 15)

This can be understood as an item X group interaction term in the ANOVA model. The conditional item bias definition proposed was as follows:

> if one holds constant the score on the underlying trait being measured, an item is biased if the probability of a right answer differs among the groups under study. (p. 15)

It should be noted that the term conditional refers to the fact that bias is defined conditional on ability level.

There are several methods for detecting item bias. The unconditional methods include ANOVA and the transformed item difficulties (TID) approach. The conditional item bias detection procedures are the item characteristic curve methods, the chi-square, log-linear and logit approaches, the regression bias method and the partial correlation index. These methods were discussed in the report and evidence concerning their effectiveness evaluated.

The three parameter item characteristic curve (ICC) procedure was reported to be the best method for detecting item bias, followed by the log-linear models and chi-square method. The TID is sometimes considered to be a good method and there is agreement that it is the best of the unconditional approaches. The conditional methods are generally more effective at item bias detection than the unconditional techniques.

Taylor argued that the chances of making errors can be minimised if item bias detection is performed using more than one sample of data and more than one method. In the document it is recommended that when item bias is to be detected a multi-method multi-sample strategy should be adopted. Furthermore, when such a strategy is applied, it is preferable that the methods and samples should be as different as possible. For example the samples could come from different geographical locations, and one conditional and one unconditional method of item bias detection used as each is based on a different theoretical approach. For a given test, the final set of biased items would be those which were identified as biased in at least half of the bias analyses using different methods and different samples.

As noted above there are several conditional and unconditional item bias detection methods. Although the three parameter ICC is the best the requirement of large sample sizes precludes the use of this method in many situations. Consequently Taylor recommended that the iterative logit method (based on the log-linear approach)

be used as the conditional technique. The TID is the best of the unconditional item bias detection procedures. It is easy to apply and can be used together with the iterative logit method.

In the report it is suggested that in addition to item bias results, other more general test information is required when the applicant pool is multicultural. In order to make decisions on the suitability of a test for different cultures, means, standard deviations, skewness, reliabilities, proportions of individuals attempting each item, item difficulty values and item-total correlations should be examined. Finally, in addition to conducting statistical item bias studies, subjective inspection of the items for perceived culture bias should be performed.

With regard to the above discussion of item bias, it should be noted that even if item bias is absent from a test, other forms of bias may be present; for example, pervasive bias, i.e. all the test items may be biased for one group. (If every item is biased for a particular group this will not be revealed in item bias analyses.)

In the report predictive bias was defined as follows:

> A test is a biased predictor if there is a statistically significant difference between the major and minor groups in the slopes, or in the intercepts, or in the standard error of estimates of the regression lines of the two groups when these regression parameters are derived from the estimated true scores of persons within each group. (p 17)

In order for a test to be examined for predictive bias, the test scores must be used to predict criterion scores. Predictive bias differs from item bias in that a criterion measure is necessary. Job performance ratings or training course marks are commonly used as criteria. Because a criterion is necessary in predictive bias research, no single predictive bias study can enable one to conclude whether a test is predictively biased or not. Even if no bias is found with respect to one criterion, bias may occur with regard to another.

In a predictive bias study it is usually assumed that the criterion is unbiased. The truth of this assumption should be considered because it is obviously sometimes incorrect. It is important that the criteria for the two groups that are being compared be the same.

Taylor included a final chapter entitled "Going beyond a priori groups" in the document. Although bias research is normally performed using groups defined beforehand, for example race and gender groups, it is suggested that more interesting information about how people answer test items could be obtained if the tests were administered and groups subsequently defined according to test responses (a posteriori groups). That is, groups are formed on the basis of patterns of item responses rather than on biographical factors such as race or gender. Taylor commented that initially such research is likely to be of theoretical interest only, but that practical applications could emerge later.

2.3 Apprentice selection: An HSRC/NTB survey of policies and methods used in the RSA with an emphasis on psychometric testing — P Holburn

2.3.1 Introduction

Because there are often many more applicants for apprenticeships than there are places available, and because it is necessary that the company select people who will contribute in a positive way to the organisation, selection procedures are applied to decide who shall be offered employment. Apprentice selection is particularly important for employers due to the large sums of money spent on apprentice training.

2.3.2 Objective of the study

The aim of this research was to survey companies in South Africa with regard to their apprentice selection methods, particularly psychometric testing. Special attention was paid to the policies and methods used for different racial and gender groups.

2.3.3 Methodology

A questionnaire covering a wide variety of topics relevant to apprentice selection was mailed to approximately 3 800 organisations employing apprentices. These organisations were of various sizes and from many sectors of industry. Six hundred and forty-four of these organisations returned completed questionnaires. The survey was undertaken during the period November 1987 to January 1988.

2.3.4 Conclusions

Many different recruitment methods were used by the various apprentice employers. The most common methods used to recruit apprentices were word-of-mouth, unsolicited applications (or walk-ins) and advertisements in newspapers. Larger organisations were more likely to make use of advertisements in newspapers, whereas smaller organisations preferred the informal methods of word-of-mouth and walk-ins. Large companies often reported making use of several methods. The method most strongly associated with the employment of a greater proportion of Asian, black and coloured apprentices was promotion from within the company.

Several problems with regard to recruitment were mentioned. These include insufficient applicants meeting the entrance requirements and the perception of the trades as dirty manual labour.

Factors which were noted to be important for ensuring fair recruitment include advertising that jobs are open to all applicants irrespective of race or gender, organisational involvement in school vocational guidance programmes and the promotion of semi-skilled employees from within the organisation.

With regard to admission requirements most organisations considered minimum schooling level and age in their preselection. Area of residence was also sometimes reported as being of importance. Some companies focused on the applicant's race and gender, which could be either to the detriment or the advantage of the applicant.

Larger companies reported being concerned with psychometric test scores as well as age and schooling level, and sometimes applicants were required to undergo medical examinations.

When the entrance requirements for the more technologically advanced trades were scrutinised, it was noted that many organisations insisted on a minimum requirement of Standard 10 with Maths and Science, particularly for black applicants. Over a third of the respondents reported that they had recently raised entry requirements because of the recession (i.e. because they could select from better qualified applicants), because they wished to indenture more capable applicants and because of technological advancement.

Of all the methods used for apprentice selection, interviews were the most frequently utilised, followed by application forms, psychometric testing and medical examinations. These last three methods were more likely to be used by the larger companies.

Almost all respondents said they make use of an interview when selecting apprentices, and for the smaller companies this is often the only method of selection. Interviews are conducted on a one-to-one or panel basis, with large companies more likely to make use of a panel. Most interviews are unstructured.

As was noted by Watt (section 2.1) screening of applicants usually occurs on the basis of information in the application forms. Organisations tend to develop their own application forms, but similar information is generally sought from applicants. The kind of information typically requested relates to highest school standard passed, age, area of residence, parents' occupation, past employment and work experience, interests and hobbies. The respondents indicated that schooling level, school marks and post-school qualifications were most strongly predictive of training success, whereas job performance was considered to be best predicted by employment history (this is a problem because most apprentice applicants are school leavers with no previous job experience). Most companies indicated that biographical information predicted well for all race groups. Although a few companies stated that they had considered the validity

of biographical information, most had not. Thus, there is a need to examine the validity of this type of data.

Around 60% of respondents reported using psychometric tests and about a third of these (notably the larger employers) performed the testing within their company. Of the companies conducting their own testing, the tests most frequently used were (in order of frequency): the Mechanical Comprehension Test, Blox, the Mental Alertness (from the Intermediate Battery) and the High Level Figure Classification Test. Personality measures were seldom mentioned and interest questionnaires did not appear to be used at all. The same tests tended to be used to select applicants from all race groups and most organisations reported using the same time limits for applicants from all races. Some organisations used different cut-offs or norms for different race groups. Smaller companies that did not have the facilities to conduct testing within their company sent apprentice applicants to other organisations where they could be tested. The Department of Manpower performed the testing for most of the smaller companies.

When organisations are concerned with fair selection and test bias, conducting validity studies should be a high priority. It makes little sense for tests to be used if they do not work well, i.e. do not predict training and job success. Around two-thirds of the respondents reported that they conduct some type of validity study (sometimes a very informal analysis).

Work sample testing, trainability testing and psychomotor testing did not appear to be frequently conducted. With reference to work sample, trainability and psychomotor testing, it should be noted that one of the reasons that these methods are seldom used is that the tests need to be developed by each individual company because standardised versions are not available. In addition many employers know little about these methods of testing job applicants.

In the report it was recommended that an apprentice selection strategy should include psychometric testing, trainability or work sample testing and possibly some

- psychomotor testing. This would enable selectors to assess the cognitive and the "hands-on" motor skills of the applicants.

The vast majority of the apprentices employed by the respondent organisations were white (84%) and in the more technologically advanced electrical and electronics trades almost all apprentices were white. The trades with the greatest number of black apprentices were the bricklaying, painting, tiling and plastering, carpentry and woodwork trades. Several companies mentioned that they were concerned with encouraging the employment of apprentices of all races. Organisations that had set targets for black appointments were more likely to have a greater percentage of black apprentices in their company.

From the data reported, virtually all of the apprentices were male, with the exception of the hairdressing industry. At the time the survey was conducted it was noted that there appeared to be very little interest on the part of employers to encourage women to enter the trades traditionally occupied by men. Many of the companies had cut back on apprentice selection because of the recession, a factor which was likely to preclude the selection of many women and blacks.

Most of the employers seemed to favour a policy of "equal" or "fair" employment. There is no exact definition of these terms and the specific policies of the companies were not ascertained. Two-thirds of the respondents said that they were in favour of the introduction of pre-apprenticeship training for people from disadvantaged backgrounds and a few indicated that they had bridging programmes for applicants from disadvantaged backgrounds.

2.4 **Selection decisions: The quest for fairness** **P Holburn**

2.4.1 Introduction

Whenever the number of job applicants is greater than the number of available

positions, selection has to occur. Some procedure is necessary to decide who shall be offered employment. Fairness in selection relates to the decisions made during selection.

Although many employers believe that the fairest approach is to select the job applicant possessing the skills and abilities that most closely match the job requirements, for many years the race or sex of an applicant has entered into selection decision-making. In the past, legislation prohibited specific racial and gender groups from performing certain jobs. Thus, race and sex, as well as ability and experience, were factors influencing selection decisions. These factors are still involved today when selection is undertaken. For some organisations fair selection entails choosing the most qualified candidate in terms of ability, whereas for others some sort of compensation for past unfair practices is perceived as necessary.

2.4.2 Objective of the study

The aim of this report was to identify the issues involved in fair selection, to discuss the philosophies and models of fair selection and to present practical guidelines for implementing fair selection in the organisation. The intention in this study was to describe various approaches to fair selection rather than to prescribe a particular method and in this way to serve as a source of information for decision makers.

2.4.3 Methodology

A literature study of fair selection was undertaken. The three philosophies of fairness and eight fairness models were described in the report. Two simulation studies illustrating the consequences of the application of several different methods of fair selection were also presented.

2.4.4 Conclusions

The three philosophies and eight models of fair selection were discussed as part of

the theoretical side of fair selection.

The three philosophies of fair selection are unqualified individualism, qualified individualism and quotas. Unqualified individualism is concerned with the maximisation of expected job performance. According to the philosophy of unqualified individualism the variables which best predict job performance for each group should be used during the selection process and the applicants with the highest scores accepted. The qualified individualism approach is much the same except that variables such as race, gender or ethnicity may not be taken into account when selecting applicants because to include such variables when making selection decisions is viewed as unfair discrimination. Quotas are concerned with the percentage of applicants selected from predesignated groups.

Although these are the three main theoretical approaches to fair selection, in practice fair selection is usually carried out according to one of two views - unqualified individualism, which can also be construed as maximising job performance, or quotas.

All of the models of fair selection discussed in this report fell into one of the above three philosophies of selection. The fairness models discussed in this report were: Cleary/Regression Model; Equal Risk Model; Thorndike/Constant Ratio Model; Subjective Regression Model; Cole/Conditional Probability Model; Equal Probability Model; Decision Theoretic Models; and Proportional Representation Models. For some of these models job performance is the most important (or only) criterion of concern, for others, other factors, for example remedying past discrimination, or obtaining more employees from groups poorly represented in the organisation, are of importance. The various models were compared with one another and the advantages and disadvantages of each noted. In particular the arguments in favour of and against the job performance maximisation/unqualified individualism (Cleary/Regression and Equal Risk Models) and quota approaches (Thorndike/Constant Ratio, Cole/Conditional Probability, Equal Probability, Subjective Regression and Proportional Representation models) were outlined.

Two simulation studies were presented which depicted the effects of a few of the models on productivity and on black and white employment figures. The simulated situation was one in which most jobs were initially held by whites.

As was expected, application of the models concerned only with job performance resulted in the highest productivity and smallest increase in black employment. Conversely, black employment increased the most with the lowest productivity, when a quota model was utilised. With this method the percentage of black applicants selected equalled the percentage of white applicants selected.

One of the most interesting findings concerned the Thorndike/Constant Ratio model, which is concerned with job performance as well as fairness to groups. When this model was applied it was found that productivity was often almost as good as that for models concerned with job performance only, while the increase in black employment was fairly substantial, particularly in the long term. It was found that unlike the Cleary/Regression Model (selection top-down purely on predicted job performance), after the simulation of 20 years, there was an 87% increase in black employment with only 0,3% lower productivity for the Thorndike/Constant Ratio model.

The explanation of this finding becomes clear if the rationale for the Constant Ratio model is examined. Selection according to the Constant Ratio Model means that the ratio of the proportion selected to the proportion successful should be the same for each group. Selection is based on accepting applicants who are expected to perform well on the job while cognisance is paid to the group to which the individual belongs. If one group of applicants scores, on average, below another group on the test and the test has validity less than one, more members from the lower scoring group will be rejected when they could have performed successfully on the job. This is because the average differences between the groups will be larger on the test than on the job.

One final comment on the simulation studies is that the consequences of the various models depend on the validity of the selection methods and the selection ratio, and the short term findings for a particular model can vary a great deal from the long term

results.

Some of the more practical aspects of fair selection were presented towards the end of the document. Guidelines for implementing fair selection in a company were outlined in point form. In this report it was emphasised that there is no single correct method of fair selection. Whether selection is perceived as fair or not depends on the values held by the various parties involved in the selection process. Thus, what is fair under one model may not be fair under another. It was also noted that some models are statistical models, whereas others are based on policy decisions. However, although statistical models of fair selection exist, the whole issue of fairness is not a statistical one, but one of policy.

The important practical points mentioned in the report were the following.

- Fair selection begins with fair recruitment. Institutions should decide explicitly who they wish to recruit and how they intend to accomplish this.

- Companies need to determine practical, workable policies based on individual conceptions of fair selection. Fair selection procedures should be practical in terms of the effort, expense and expertise required. Policies should be more specific than statements such as "we are an equal employment company". In fact, the organisation should ideally produce a written fairness policy. The results of selection decisions based on such a policy should be continually noted and the policy revised and updated when necessary.

- Many factors should be kept in mind when a policy document is drawn up. These include the needs, interests and power of various interested parties, the future supply of labour, the projected growth of the selecting institution, the future manpower requirements, Government policy, legislation and legal precedents. The institution will also have to decide if they want to consider defining groups or if selection is to be based purely on individual merit. If selection is to focus on groups there must be some way in which these will be

differentiated, and if selection policies involve quotas, the selecting institution should consider the possibility of charges of reverse discrimination.

- Whereas in the past the main concern of the selecting institution was maximising job performance other factors can be considered when employing personnel, for example maximising cultural diversity in the work place. (This is particularly important in the light of past discriminatory practices.)

- In attempts to implement fair selection policies, internal training courses can be designed and bridging and other educational programmes implemented. It is important to be aware that people can be trained and jobs redesigned.

- An organisation that is concerned with fairness should examine the validity (preferably criterion-related validity) of their selection instruments for different groups. The use of selection instruments which are not valid for all groups are likely to result in contentious decisions, particularly if greater proportions of individuals from some groups than others are rejected on the basis of performance on these instruments.

- The information obtained from selection should be treated confidentially.

- Job analyses should be conducted so that the job skills required are clearly understood and appropriate selection instruments chosen.

- Once an organisation has formulated a fair selection policy it must be communicated to all employees, particularly those involved with selections. Staff need to be informed of any new policies and procedures.

- Probably one of the most important points with regard to fair selection is that selection must be **perceived** as fair by all parties.

2.5 Test bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox and High Level Figure Classification tests

P Holburn

2.5.1 Introduction

Many organisations in South Africa make selection decisions on the basis of psychometric test scores. Recently, however, several people concerned with testing have spoken of the need to investigate the suitability of tests for various race groups.

The first studies investigating test bias in South Africa were undertaken by researchers at the Human Sciences Research Council in the late 1980s. Although these initial studies yielded interesting and informative results, there has been very little such research conducted, and the focus has been on coloured-white comparisons on educational tests. Not only is there a need for more test bias research, but studies investigating test bias between black and white examinees in industrial situations is urgently needed.

2.5.2 Objective of the study

The objective of the report was to empirically examine bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox and High Level Figure Classification tests, which were noted in section 2.3 to be frequently used for apprentice selection and to report the results.

2.5.3 Methodology

Item bias and predictive bias analyses were undertaken. The item bias detection procedure adopted was the one outlined by Taylor (1987) and reported in section 2.2. A multi-method multi-sample strategy was followed. The transformed item difficulty (TID) and iterative logit methods were applied to the test scores of Asian, black, coloured and white apprentices from three South African companies.

2.5.4 Conclusions

Mean differences were observed between the various race groups on all the tests. The black-white mean difference was the largest and was usually greater than one standard deviation. Consequently the TID method was not able to differentiate between difficult and highly discriminating items and the results of the TID method could not be used.

The Mechanical Comprehension test was found on the basis of internal reliability consistency figures (KR20) to be unsuitable for multicultural apprentice selection and it was recommended that this test should not be administered.

The Intermediate Mental Alertness test had a fairly low internal consistency reliability figure for the black sample, but seemed acceptable for the other groups. Seven items emerged as biased against the black applicants and it was recommended that a correction factor be applied to the test scores of the black applicants to deal with the bias. For example, if the test has 30 items and 10 items are biased against the black group, the test should be scored out of 20, the number of unbiased items, for the black group. The total score out of 20 is then multiplied by 30 and divided by 20 to equate the scores with the other groups that have completed all 30 items. Alternatively, one could set different cut-offs for the black group. The bias against the Asian, coloured and white groups seemed minimal; at most one item was biased.

The High Level Figure Classification test appeared to be suitable for all the examinees, with at most one item biased against any group. For all applicants the internal consistency reliability figures were excellent.

High internal consistency reliability figures for all groups were obtained with the Blox test, although some bias did emerge against black and coloured applicants. Three items were biased against each of these groups. As with the Intermediate Mental Alertness test, it was recommended that the scores of examinees from groups where bias was present be corrected.

Predictive bias analyses were undertaken for the full 24 item High Level Figure Classification test and the 51 item Blox test for a sample of black, coloured and white apprentices from the Cape.

When the High Level Figure Classification test and the Blox test were used to predict two criteria of first-year apprentice performance on training modules predictive bias was noted for the Blox test between the black and white apprentices. The black apprentices obtained a lower average test score but a higher average job performance score than the white apprentices, and thus the bias was to the detriment of the black sample. For the Blox test, the black apprentices had an average test score 3 points below that of the white apprentices, whereas their average performance score was around 4% higher. However, it was noted in this report that for the predictive bias research the samples were small; hence, the results should be considered to be tentative.

In a country where multicultural selection is performed and the effectiveness of tests for different cultures is contentious and not well understood, employers of apprentices should consider the possible presence of predictive bias in tests. When predictive bias is present, the regression lines indicating the linear relationship between the test scores and criterion scores (apprentice performance) are not the same for the different groups. That is, similar test scores will predict different criterion scores. For example, in the predictive bias research reported in this document, the black apprentices had lower average test scores but higher average criterion scores than white apprentices for one of the tests.

Although selectors may decide on the basis of the evidence presented in this report that tests should not be used for apprentice selection because they contain bias, it should also be remembered that other selection instruments may be biased. It is also necessary that selectors bear in mind that when bias occurs in selection procedures it is easier to ascertain and eradicate this bias in tests than in any other method.

In the conclusion of the report it was pointed out that when bias is found it is possible

to calculate correction factors based on the amount of bias identified in the test. However, this procedure can probably be performed only by specialised personnel, for example the supervising psychologist responsible for the control of tests in the organisation. Other alternatives that employers could consider include making use of separate norms or the method of "banding" (see section 3.2.2 point 9). It was noted in this report that the method of dealing with test bias that is chosen by the company will depend on their individual conception of the fair use of tests.

3. CONCLUSIONS AND RECOMMENDATIONS

In this final chapter recommendations for fair and unbiased apprentice selection are made. Because different parties will have different ideas about what constitutes fair selection, a detailed outline for apprentice selection will not be presented. Instead, guidelines will be drawn and the reasons for these put forward.

3.1 Factors to be taken into account when making selection decisions

3.1.1. Selection occurs firstly because there are too many applicants for the available positions and secondly because organisations wish to ensure that the individuals selected will be capable of performing effectively. Thus, selection methods should discriminate among applicants with the results that some will be accepted and others rejected. The basis on which selection or rejection is to occur must be clearly decided on. Is it to be merit, or some other criterion?

3.1.2 Many selection methods test the abilities and knowledge that an applicant has acquired in the past. However, not everyone will have had the same opportunities to acquire the same knowledge because different applicants will have had different learning experiences. Although this is evident with regard to groups, for example racial and gender groups, this principle applies on an individual level as well. For example, if two white male candidates are considered, one may have had more opportunities to acquire knowledge which would be of benefit in selection (attended a better school; worked on cars in his father's garage over the weekend). Before considering any fair selection policy and implementing any selection procedure it must be recognised that an applicant who does not meet certain selection criteria may be capable of doing so should appropriate opportunities be provided.

3.1.3 Apprentices are selected in order to be trained. Selection decisions are crucial as they can be costly should the apprentice fail to complete his or her

course successfully. This also means that consideration can be given to potential. Unlike selection for certain jobs, the candidate does not have to be able to effectively perform artisan work on acceptance.

3.1.4 Although it is necessary that apprentices pass their training courses and modules, the main concern should be with performance on the job, i.e. artisan work. Because much of the apprentice's training is similar to the work the artisan performs, it is likely that a successful apprentice will become a highly competent artisan.

3.1.5 The choice of selection procedure depends on the type of job. A nationwide survey (discussed in section 2.2) revealed that technological developments had resulted in considerable changes to some artisan jobs. This in turn means that the selection requirements should be specified accordingly. In trades strongly affected by technological advancement more weight may be given to cognitive abilities because frequently less manual work is performed and the artisan's job is sometimes fairly similar to that of the technician. Greater attention may be given to selecting applicants who have the learning ability to update their knowledge and skills when necessary.

3.1.6 The size of the company also affects selection procedures. The survey demonstrated that small companies differ from larger ones in their choice and application of selection methods. Small employers tend to have fairly informal selection methods, often making use of just an interview and sometimes a work sample test. They are also less likely to be accused of unfair or biased selection practices. However, this does not mean selection should be unfair.

3.1.7 In order for a company to decide which selection methods to use and how to use them, the characteristics of the job must be known to the selectors (Taylor, in press). In other words, a job analysis must be conducted. Many different methods exist for performing a job analysis; some entail the

completion of lengthy and complex forms, whereas others are written paragraphs of the requirements of the job. However, these methods all enable the employer to ascertain which skills are required by the job incumbent. An example of a job analysis method which is available and used in South Africa is the PAQ (position analysis questionnaire) (McCormick, Jeanneret & Mecham, 1980). Once the selectors know what the job requirements are it is easier to decide if selection procedures are relevant to the job. Although for some employers fairness will incorporate more than just selecting those expected to perform best on the job, many fair selection policies start with the use of methods that are designed to select apprentices likely to perform best.

3.1.8 The use of selection instruments which result in large proportions of a particular group being rejected is contentious. When this occurs it is particularly important that the validity of the selection instruments be proven to be high for all applicants. When selection methods result in many applicants from one group being rejected, alternative selection methods which do not have this effect should be considered.

Validation studies should be performed separately for all cultural groups (gender as well as racial groups). Methods with a high degree of face validity are more likely to be approved of and perceived of as fair by candidates.

3.2 **Guidelines for implementing fair selection**

3.2.1 *Selection methods*

Different selection methods will first be considered separately before the selection procedure as a whole is discussed.

3.2.1.1 Recruitment

For some organisations it may be difficult to see how recruitment could be unfair because any applicant can apply for an apprentice position. However, for other companies fair apprentice recruitment may entail a proactive attempt to obtain applications from groups significantly underrepresented in the company.

Because most of the current apprentice workforce is white and male, recruitment methods should ideally be proactive in order to increase the representation of other groups in the company. For this to happen informal recruitment methods, for example the use of social networks such as family and friends, "word of mouth" or "walk ins", should be used minimally, if at all.

It is desirable that recruitment methods should reach all racial and gender groups. Advertising can be undertaken within and outside the company and be targeted at all possible applicant groups. Schools and technical colleges can be visited in order to encourage future applicants. It should always be made clear that apprenticeships are open to all races and both sexes. Prospective applicants are more likely to apply for positions for which they know they are eligible. A company is also more likely to attract high level candidates from all groups if it's employment practices are perceived as fair.

A recession usually results in less effort to recruit, which can be a problem as attempts to increase the representation of different cultural groups in the organisation are curtailed. However, internal recruitment could be implemented in its place. Existing employees can be retrained and positions upgraded. People who have already been working in the company for some time may have demonstrated potential for promotion.

3.2.1.2 Application forms

Application forms are frequently used as a prescreening device and are fairly effective for this purpose. A prescreening measure is often necessary for apprentice selection

because many applicants may be available, particularly in recessionary times, and some of these may be totally unsuitable.

The information requested on application forms should be relevant to apprentice performance. Information which is of a highly personal nature or is possibly unfairly discriminatory should not be requested unless it can be shown that it is related to apprentice performance. This implies that the validity of application forms should be ascertained.

Although a careful examination of items can result in the elimination of many unsuitable questions, it is preferable to evaluate the importance of biographical information for apprentice success using data from one's own organisation. Because many employers reject applicants on the basis of biographical information it is particularly important that the predictive validity of these forms be evaluated.

Entry level requirements should be scrutinised. They may be unnecessarily high resulting in a greater percentage of applicants from one group being rejected. Highest school standard passed is an important piece of biographical information which employers seek. Because this information can unfairly prejudice the chances of many black applicants, schooling requirements should not be set too high unless they are absolutely necessary for the apprentice to succeed in training. Selecting candidates whose family members are artisans in a similar or the same trade is also unfairly discriminatory.

3.2.1.3 Interviews.

Almost every employer interviews prospective apprentices. Although interviews have been found to have low validity (Farrell, 1986; McCormick & Ilgen, 1987), it is unlikely that companies will do away with the interview. In the likely event that the interview is retained, the fairness as well as validity of this selection method can be improved by following certain principles.

Structured interviews that include job-related questions are considered the most valid and fair (Campion, Pursell & Brown, 1988; Pursell, Campion & Gaylord, 1980). A structured interview entails asking all applicants a fixed set of questions for a particular job. The questions should be job-related and should aim to elicit responses which reveal how well the candidate can be expected to perform as an apprentice. A job analysis can be used to facilitate the formulation of appropriate questions. Questions which ask for personal information, and information which is possibly discriminatory against certain race or gender groups, should not be asked unless it is deemed essential for effective apprentice performance.

Interviews are usually performed towards the end of the selection process and can often be used to clear up any problems or uncertainties that have arisen earlier.

When companies interview apprentices the ideal is a standard interview format with set questions which ask for information relevant to apprentice success. Small companies that use the interview as the only method for selecting apprentice should follow these guidelines. When a structured format is used, it is possible to validate the interview. If an organisation has shown that their interview has high validity, it is difficult to argue successfully that the results of an interview are biased.

3.2.1.4 Psychometric testing

Because this research project was predominantly concerned with an investigation of cultural bias in tests this method will be covered in greater depth than the others.

It was apparent from the survey that many employers, particularly the large companies, conduct psychometric testing. A major problem with psychometric testing is that very little information has thus far been available on the meaning of test scores for black examinees.

Test bias studies were conducted on the four tests most frequently used for apprentice selection as identified in the survey (section 2.3). These tests were the Mechanical

Comprehension test, the Blox test, the High Level Figure Classification test and the Intermediate Mental Alertness test.

It was recommended that the Mechanical Comprehension test A/3/1 not be used for multicultural apprentice selection because this test has extremely low reliability for blacks. This finding was not unusual and it has been observed and commented on by many tester users in practice. Another problem with this test is that it is old, and many of the newer technological aspects are not examined. Although the Mechanical Comprehension test was the most frequently used test for apprentice selection, the test is not suitable and employers should not include this test in their batteries. Because there does seem to be a need for a mechanical test the HSRC is at present completing the development of a mechanical test battery which could be used for this purpose. A practical work sample test could also be developed by employers if they wish to test the mechanical skills of applicants.

The remaining three tests were subjected to item bias studies. In all cases item bias emerged. This indicates that the test scores do not mean the same thing for the different race groups. Most of the bias was against blacks; little if any bias was found against coloureds; and no bias appeared when the Asian and white samples were compared.

The Intermediate Mental Alertness test had seven items biased against and one item biased in favour of the black applicants. The biased items were mostly alphabetic type items, i.e. alphabetic codes and alphabetic series. Because the biased items did not seem to be absolutely necessary for successful apprentice performance it seems appropriate that some correction be made to the test scores of the black applicants (see section 2.5.4).

The Blox test had three items biased against blacks and three against coloureds. The method for dealing with bias in the Intermediate Mental Alertness discussed in section 2.5.4 can be applied for the Blox as well. The High Level Figure Classification test had only one item biased against black applicants (and one item biased in favour of

coloured candidates) and it appeared that this test could remain very much as it is for apprentice selection.

Whereas the High Level Figure Classification test is a nonverbal general reasoning test, the Intermediate Mental Alertness test is a verbal test of general reasoning ability. Because apprentices do not need to be highly verbally skilled, greater emphasis should be placed on the candidates' High Level Figure Classification test scores. If an applicant has a high score on the High Level Figure Classification test but a low score on the Intermediate Mental Alertness, verbal ability could well be the reason as both tests are assessing general reasoning ability.

Despite the bias that emerged, at this stage these three tests (Intermediate Mental Alertness, Blox and High Level Figure Classification tests), together with the new mechanical battery, are some of the HSRC tests recommended for apprentice selection.

Item bias research is concerned with the meaning of tests and test scores. Of particular practical importance, however, is how well the tests predict apprentice performance.

Although the predictive bias analyses that were conducted produced only tentative results, it seemed as though some predictive bias was present against the black group for the Blox test. Compared to the white group, black apprentices obtained, on average, lower test but higher apprentice performance scores.

When predictive bias is found, similar test scores are predicting different criterion scores for the different groups. In this research the test scores of black examinees underpredicted their apprentice performance scores. In this situation it is evident that there is bias against the group and some correction should be made.

Predictive bias studies should be conducted in each organisation to determine if predictive bias is present. If predictive bias is found then the regression lines of test

on criterion can be drawn for each group. A level of criterion performance that is considered satisfactory can be determined and the test cut-off scores corresponding to the criterion cut-off point set for each group.

To conclude this section on psychometric testing it should be noted that several general principles should be followed when multicultural apprentice testing is undertaken (these also pertain to testing in general).

Firstly, the test must be of the appropriate difficulty level. If a test is clearly much too difficult or too easy for a particular group, it should be replaced with a more suitable version.

Secondly, if the reliability values are too low for a group then the test should not be used. Similarly, there is little point in taking the trouble and time to administer a test that is not valid.

Thirdly, predictive bias studies should be undertaken. Employers can also consider performing item bias research.

Fourthly, testing must be conducted in accordance with the legal and professional requirements.

3.2.1.5 Psychomotor testing

Few employers reported making use of this technique for apprentice selection, although some expressed an interest should standardised versions become available.

Because apprentice and artisan work has a "hands-on" component, psychomotor tests could be useful in apprentice selection. It is therefore recommended that some standardised psychomotor tests be developed, for example a general eye-hand coordination test or a two-hand coordination test. It is also entirely feasible for the large employers or the industry bodies to design their own psychomotor tests.

3.2.1.6 Trainability testing

From the survey of apprentice selection procedures, a number of employers reported that they already made use of trainability tests. The main stumbling block preventing more people from using them is probably the absence of standardised trainability tests. Here as well, employers could develop their own instruments.

Trainability tests have often been found to be the selection instruments with the highest validity. They are also perceived as being the fairest of all selection methods because they are obviously related to the job. Therefore they can help to fulfil the need for effective and fair selection methods.

A trainability test occurs in several stages. An instructor first demonstrates to the examinees how to perform several tasks (which should be similar to tasks performed on the job, but broad and general enough to cover most of the job demands). The examinees are subsequently required to perform the same tasks on their own. Their performance is rated and a score assigned to each person.

Trainability testing is particularly appropriate for apprentice selection because it provides an indication of how well applicants can learn skills. As was mentioned earlier, apprentices are selected in order to undergo training. Therefore it is not necessary to select people who will have to perform efficiently immediately. Trainability testing is discussed in more detail by Taylor (1982) and Wheeler (1989).

Conceptions of fairness differ among individuals. Some people may regard selecting the best people for the job as fair. Others may take into account the fact that applicants will have had different opportunities to develop their skills and abilities. Trainability testing is one way of addressing both issues. Although previous experience can be an advantage when completing a trainability test, a trainability test is a type of learning potential assessment device in that examinees are afforded some opportunity to show how they benefit from instruction.

3.2.2 *The selection procedure*

It is essential that the above selection methods be combined into a selection procedure in order for apprentice selection to be conducted. Several selection methods should be used as it is possible to obtain a more accurate picture of the likely success of a candidate in this way. Quantitative methods have an advantage in that they are less open to abuse and bias can be more easily detected.

The following points should be borne in mind when a selection procedure is drawn up.

1. Recruitment should be proactive and targeted at all people who are eligible for apprenticeships. In this way it is possible to improve the suitability of applicants.

2. Because the number of apprentice applicants can be very large, application forms are useful and necessary to eliminate obviously unsuitable candidates. The information that is requested from application forms must be related to successful job performance, particularly if a high percentage of one group is rejected and of another is selected. Employers of apprentices should consider validating their application forms and industry bodies could be involved in developing a standard application form.

3. Psychometric testing can be a useful adjunct to application forms provided that possible bias in the tests is considered. Predictive bias studies can be conducted by those who use psychometric tests, and when predictive bias is found appropriate action can be taken by the personnel in charge of the use of tests. Depending on the organisation's conception of fairness, the degree of test bias may be calculated and scores corrected. Separate norms or test score bands could also be used or the organisation may prefer to leave test scores as they are and pay more attention to other selection methods. These different uses of test scores are discussed in more detail under point 9.

4. Psychometric tests can be used to screen out applicants who clearly do not have sufficient cognitive ability to succeed. Because there seems to be bias present for apprentices, and cognitive ability although necessary is not sufficient, selection should not be based solely on psychometric test scores. A test score can be set which indicates the minimum level at which apprentices can be expected to succeed. Those who obtain a score higher than this cut-off score can then be allowed to proceed through to the next stage. Because some trades are more technologically advanced than others, it may be necessary to set higher test score cut-offs for applicants for these positions.

5. When psychometric tests are used for multicultural selection a nonverbal test should be included. With applicants whose home language is not the language in which they were tested, greater attention should be paid to the results of nonverbal tests. Companies could include psychomotor tests in their test battery as these can be useful for apprentice selection.

6. Trainability testing is strongly recommended as the most useful method of multicultural apprentice selection. It is recommended that trainability testing follow psychometric testing. Application forms and psychometric tests can be used to identify candidates who will clearly not succeed and trainability tests can be used to select the best candidates from among those who could possibly be successful.

   Trainability tests can be based on the modules the apprentices have to pass in order to continue their training. Tasks from the various apprentice training modules can comprise the trainability test. These tasks should cover the most important skills a successful artisan requires.

7. Despite the finding that interviews do not have high validity and are easily subject to influence by interviewer prejudice, they will still invariably be used. To increase the validity of the interview and decrease the possibility of

interviewer bias, the information that is elicited should be job-related. Any question which is likely to result in a greater proportion of one group than another being rejected must elicit information relevant to job performance.

8. Many of the small companies do not conduct formal rigorous apprentice selection and are not often carefully scrutinised for instances of unfair employment practices. Nevertheless for many of the small companies, selection should still be fair, even if the procedure is not as formal. Interviews should be job-related and structured. Small employers can also use an informal work sample test, as was seen to be happening from the survey.

9. The organisation must decide on a fair selection policy. It is not possible to prescribe to an employer how to select fairly because each person has his or her own conception of fair selection. There are two main factors to consider when such a decision is made: How important is the goal of maximising job performance and how important is the goal of racial diversity in the workplace? Once the employer has considered and answered these questions, fair selection procedures can be implemented.

When the test scores for each candidate have been obtained (or alternatively a score for the entire selection procedure if this is available), the organisation can make use of these scores in a way that accords with their policy of fairness. Although the procedures put forward in points 1 through 8 are those recommended for apprentice selection, it is possible to make use of these methods in different ways depending on one's view of fairness.

(i) If job performance is the sole concern of the selecting institution, selection can occur in strict top-down order on the basis of scores. Those with the highest scores will be the first selected. The advantage of this method is that it will ensure the most productive workforce. However, because black applicants score lower on the tests than whites on average, the workforce is likely to remain primarily white. If the organisation wants to select strictly

on merit then it is necessary that the bias in selection methods be ascertained and the examinees' scores corrected for the bias.

(ii) A second method is to make use of separate norms. When this method is chosen the scores of a candidate are compared with norms from his or her own racial or gender group. Within-group percentile scores are calculated and the subsequent percentile ranks from all the different groups brought together. Selection is thus top-down on the basis of percentile scores. When this method is applied it is possible that one candidate may have a lower raw score yet a higher percentile rank than another applicant. The use of separate norms favours the lower scoring black group.

When separate norms are used the initial concern is one of attaining cultural diversity. The first goal that must be attained is a balanced representation of all groups in the workforce. Once that has been dealt with, job performance is then focused on. The advantage of this method is that all groups will probably be well represented in the organisation. The use of separate norms has been endorsed in the USA by the National Academy of Sciences (Hartigan and Wigdor, 1989) because selection top-down on the basis of test scores is not always fair to the lower scoring group. Because tests have less than perfect, and sometimes rather low, validity, errors are made in selection which work to the disadvantage of the lower scoring group. The percentage of members of the lower scoring group who are likely to be rejected when they could have performed well on the job is greater than the same percentage for the higher scoring group.

There are also certain disadvantages associated with the use of separate norms. Firstly, the aim is not to maximise productivity and it can be argued that standards may drop. Furthermore members of the lower scoring group may be labelled incompetent and possibly considered to be token appointments. They may feel they did not succeed on their own.

(iii) Due to measurement error it is possible that statistically two test scores may not differ reliably from each other. For example the two test scores 26 and 28 may not be reliably different. It is possible to calculate the difference between test scores necessary for them to be regarded as reliably different. For example the calculations may show that test scores need to be at least five points apart for them to be regarded as reliably different.

Another method of fair selection is to make use of bands[2]. When fixed bands are used in selection, the top score obtained from a particular testing session is used to form the top score in the band, i.e. the top of the band. If the band width has been calculated to be five, the bottom score in the band will be five points below the top score. All the scores in this band are not reliably different from one another and are treated as equal. If the number of applicants falling within this band is less than the number of candidates required, a band below this one can be formed in exactly the same way. When the number of candiates to be selected is less than the total number falling within a particular band, several methods can be used to choose between them. Selection could be random because all are assumed to have the same score. Alternatively, if racial or gender diversity is an important goal, all the candidates from the groups poorly represented in the company can first be selected from the band.

When bands are used, job performance and productivity are the initial concerns. However, the goal of racial or gender diversity in the workforce can also be accommodated with this method. This procedure falls somewhere between strict top-down selection on the basis of test scores and the use of separate norms. The former method is concerned only with work performance, whereas separate norms ensure workforce diversity.

---

[2]Cascio and colleagues in the USA have recently proposed the idea of bands for selection (personal communication). Although it may seem that this method is complicated, the calculations are in fact extremely simple. To calculate the band width at most four numbers need to be multiplied together. To date it has proved extremely difficult to locate any published information on bands in South Africa and anyone interested in this method should contact the HSRC in Johannesburg for more information.

When bias is present in tests but has not been corrected for, then the use of bands or separate norms for preferential selection of some groups may not constitute affirmative action. If, for example, there is bias against group A and this has not been eradicated, then the selection procedure is biased if the same cut-off scores are used for all groups. For this reason the use of bands or separate norms has been proposed as a way of offsetting bias in tests. However, unless the exact amount of bias in a test is known, it is not possible to be certain whether the application of the bands or separate norms is merely adjusting for bias, or includes a component of affirmative action. Unfortunately at this stage in South Africa we do not yet know for certain the extent of cultural bias in tests. It is thus not possible at present to calculate exactly how much the use of separate norms or bands will offset the degree of bias in tests.

However, it seems that the use of separate norms would totally offset the presence of bias in tests and may well include some affirmative action in favour of the lower scoring groups. Bands, particularly if there is a large amount of bias present in tests, may not offset all of the bias. Therefore bands do not offer as great an advantage to the lower scoring groups as do separate norms.

In addition to performing selection according to one of the above methods, a company may also decide that fairness entails spending more resources on providing bridging courses for unskilled or semi-skilled black employees. These training courses may be implemented to prepare candidates for apprenticeship training.

REFERENCES

Campion, M A, Pursell, E D & Brown, B K (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41*, 25-42.

Farrell, B M (1986). Recruitment: The art and science of employment interviews. *Personnel Journal, (5)*, 91-94.

Hartigan, J A & Wigdor, A K (Eds.) (1989). *Fairness in employment testing: Validity generalisation, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press.

Holburn, P T (1989). *Apprentice selection: An HSRC/NTB survey of policies and methods used in the RSA with an emphasis on psychometric testing. C/Pers 406*. Pretoria: Human Sciences Research Council.

Holburn, P T (1991). *Selection decisions: The quest for fairness. An HSRC/NTB report. C/Pers 424*. Pretoria: Human Sciences Research Council.

Holburn, P T (in press). *Test bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox and High Level Figure Classification tests*.

Human Sciences Research Council (1985). *The HSRC/NTB investigation into the training of artisans in the RSA*. Pretoria: Human Sciences Research Council.

Hunter, J E & Hunter, R F (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

McCormick, E J & Ilgen, D R (1987). *Industrial and Organisational psychology, 8 ed.* Englewood Cliffs, N.J.: Prentice-Hall Inc.

McCormick, E J, Jeanneret, P R & Mecham, R C (1980). *Position Analysis Questionnaire*. Pretoria: Human Sciences Research Council.

Pursell E D, Campion, M A & Gaylord, S R (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal, (11)*, 907-912.

Schmidt, F L & Hunter, J E (1986). Employment testing: Old theories and new research findings. In F J Landy (Ed.) *Readings in industrial and organisational psychology*. Chicago, Illinois: The Dorsey Press.

Taylor, T R (1982). Trainability testing: Pros and cons. *South African Journal of Psychology, 12*, 51-56.

Taylor, T R (1987). *Test bias: The roles and responsibilities of test user and test publisher*. Special report Pers 424. Pretoria: Human Sciences Research Council.

Taylor, T R (in press). *Fairness in the use of information to make decisions about people in the workplace*.

Watt, J (1987). *Investigation of apprentice evaluation procedures with specific reference to test bias and fairness. Pilot study submitted to the National Training Board*. Johannesburg: National Institute for Personnel Research.

Wheeler, J J (1989). *A guide to the development of trainability tests. Report P-108*. Pretoria: Human Sciences Research Council.

ISBN 0 7969 1286 6